

Computing electricity spot price prediction intervals using quantile regression and forecast averaging

Jakub Nowotarski · Rafał Weron

Received: 31 December 2013 / Accepted: 6 August 2014 / Published online: 19 August 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract We examine possible accuracy gains from forecast averaging in the context of interval forecasts of electricity spot prices. First, we test whether constructing empirical prediction intervals (PI) from combined electricity spot price forecasts leads to better forecasts than those obtained from individual methods. Next, we propose a new method for constructing PI—Quantile Regression Averaging (QRA)—which utilizes the concept of quantile regression and a pool of point forecasts of individual (i.e. not combined) models. While the empirical PI from combined forecasts do not provide significant gains, the QRA-based PI are found to be more accurate than those of the best individual model—the smoothed nonparametric autoregressive model.

Keywords Quantile regression averaging · Prediction interval · Quantile regression · Forecasts combination · Electricity spot price

1 Introduction

Since the deregulation of electricity markets in the 1990s, electricity spot price forecasting has attracted a lot of attention. A generator, a utility company or a large industrial customer able to forecast the volatile wholesale prices with a reasonable accuracy can adjust its bidding strategy and own production or consumption schedule to maximize profits in day-ahead trading. However, in contrast to other tradable commodities, electricity is very special (Eydeland and Wolyniec 2003; Kaminski 2013;

J. Nowotarski (✉) · R. Weron
Institute of Organization and Management, Wrocław University of Technology,
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: jakub.nowotarski@gmail.com

R. Weron
e-mail: rafal.weron@pwr.edu.pl

Weron 2006). It is non-storable (economically) while power system stability requires a constant balance between production and consumption. At the same time, electricity demand is dependent on weather (temperature, wind speed, precipitation, etc.) and the intensity of business activities (working hours, weekdays vs. weekends, holidays and near-holidays, etc.). On one hand, these specific characteristics result in extreme—unobserved in any other market—spot price volatility, on the other, they have motivated intensive research efforts towards short-term forecasting of electricity prices, for a recent comprehensive review see Weron (2014).

A wide range of econometric or statistical models have been suggested in the literature, including (auto-)regressive models, (seasonal) ARIMA, AR-GARCH, jump-diffusions, factor models and regime-switching models (see e.g. Bierbrauer et al. 2007; Conejo et al. 2005; Garcia-Martos et al. 2012; Karakatsani and Bunn 2008; Kristiansen 2012; Maciejowska and Weron 2013; Misiorek et al. 2006; Weron and Misiorek 2008). However, in terms of predicting spot price movements, each model specification yields a different forecast. Despite the diversity of models it is impossible to select the most reliable one. For instance, Aggarwal et al. (2009) compared results from as many as 47 publications and concluded: *there is no systematic evidence of out-performance of one model over the other models on a consistent basis*. This fact is a good motivation for considering combining electricity spot price forecasts. Surprisingly, this approach has not been undertaken in the literature until very recently, see Bordignon et al. (2013), Nowotarski et al. (2014) and Raviv et al. (2013). All three cited papers yield similar conclusions—they support the benefits of combining forecasts for deriving more accurate and more robust point forecasts of electricity spot prices.

However, in some applications, such as risk management or bidding with a safety margin, one is more interested in predicting the variability of future price movements than simply point estimates. From a practical point of view prediction intervals (PI), and density forecasts even more so, provide additional information on the evolution of future prices. In particular, they allow for a better assessment of future uncertainty and for planning of different strategies for the range of possible outcomes indicated by the interval forecast (Chatfield 2000). While there are a variety of empirical studies on forecasting electricity spot prices, interval or density forecasts have not been investigated very extensively to date. Most notable exceptions include Bunn et al. (2013), Misiorek et al. (2006), Nogales and Conejo (2006), Weron and Misiorek (2008) and Wu et al. (2013). Yet, to our best knowledge, prediction intervals have not been considered in the context of combining electricity spot price forecasts. It should be noted, though, that the idea of combining interval forecasts is not new by itself (see e.g. Timmermann 2006; Wallis 2005).

The contribution of the article is twofold. First, we address the above mentioned unresolved issue of constructing PI from combined electricity spot price forecasts. We do this by constructing empirical PI—as in Weron and Misiorek (2008)—for two methods of forecast averaging, i.e. simple average and least absolute deviation (LAD), that proved to be robust and accurate in our recent point forecasting study (Nowotarski et al. 2014). Second, we propose a new method for constructing prediction intervals using the concept of quantile regression and a pool of point forecasts of individual (i.e. not combined) time series models. The latter can be viewed as a natural extension of LAD averaging to an arbitrary quantile.

The remainder of the paper is structured as follows. In Sect. 2, we present the electricity price data used in this study. In Sect. 3, we first present two benchmark individual models (ARX and SNARX) and two averaging schemes (SIMPLE and LAD). Then we introduce the new method—Quantile Regression Averaging (QRA)¹—for constructing PI from point forecasts of individual methods. In Sect. 4, we evaluate the forecasting performance of the five tested approaches for computing PI in terms of unconditional and conditional coverage. Finally, in Sect. 5, we wrap up the results and conclude.

2 The data

The electricity spot price data was downloaded from the GDF Suez website (www.gdfsuezenergyresources.com) and contains hourly day-ahead locational marginal prices (LPMs) for the Jersey Central Power and Light Company (JCPL) of the Pennsylvania–New Jersey–Maryland (PJM) Interconnection (U.S.). The time series of hourly air temperatures (in Fahrenheit degrees) for New York City was obtained from NOAA’s National Climatic Data Center (www.ncdc.noaa.gov).

We should note here that the spot electricity market is actually a day-ahead market that does not allow for continuous trading. This is a result of system operators requiring advance notice in order to verify that the schedule is feasible and lies within transmission constraints. In a day-ahead market agents submit their bids and offers for delivery of electricity during each hour (or half-hour in some markets) of the next day before a certain market closing time. Thus, when dealing with the forecasting of intraday prices it is important to recall that prices for all spot contracts of the next day are determined at the same time using the same available information (Bierbrauer et al. 2007; Peña 2012; Weron 2006). The system price is then calculated as the equilibrium point for the aggregated supply and demand curves and for each of the hourly (or half-hourly) intervals.

The dataset used in this study comprises hourly price and temperature data for the period August 22, 2010–January 14, 2012, see Fig. 1. The prices and temperatures were preprocessed in the standard way for ‘missing’/‘doubled’ values resulting from changes to/from the daylight saving time: the ‘missing’ values were interpolated from the two neighboring observations while the ‘doubled’ values were averaged to yield one observation for the doubled hour (as in Nowotarski et al. 2014; Weron and Misiorek 2008). The logarithms of hourly temperatures T_t (in Fahrenheit) were used as the exogenous variable in the time series models for the log-prices (no additional transformations were required since the temperatures were above 1°F in the studied period). This selection was motivated by a roughly linear dependence between these two variables. The mean log-price and the median log-temperature were removed to center the data around zero, as in Conejo et al. (2005) and Misiorek et al. (2006).

In Fig. 1 the studied dataset is split into three subsets. The first date—August 22, 2010—marks the start of the calibration period for the 12 individual models (for model definitions see Sect. 3.1). The first prediction of these models is made for February 11,

¹ Matlab code is available from <http://ideas.repec.org/s/wuu/hocode.html>.

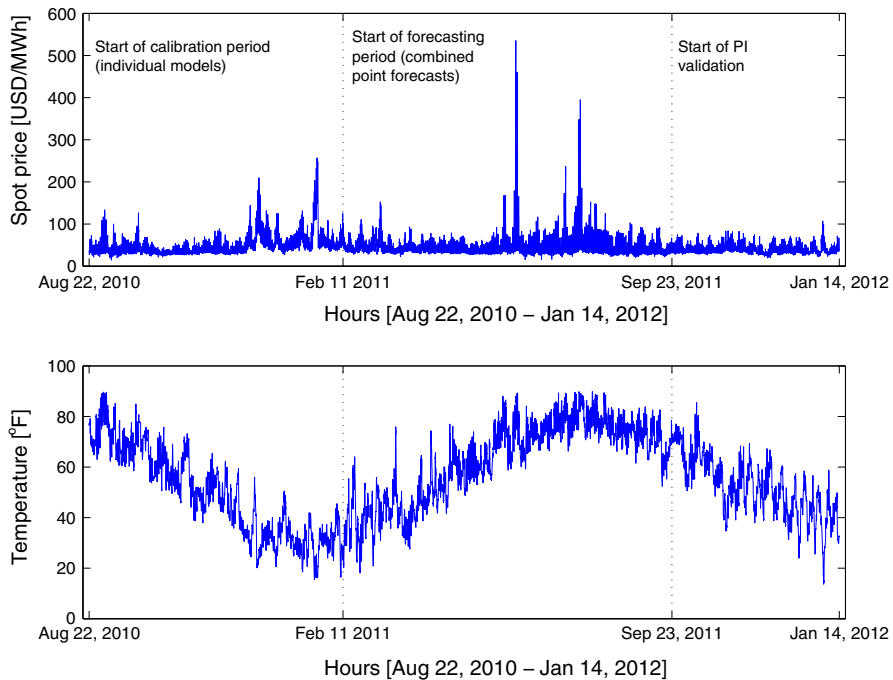


Fig. 1 JCPL (PJM) hourly system prices (in USD/MWh) and hourly air temperatures (in Fahrenheit) for the period August 22, 2010–January 14, 2012. The *first vertical dotted line* marks the start of the forecasting period for individual models, the *second*—the *beginning of the validation window* for obtained prediction intervals

2011 (the second date). Then the window is expanded by one day, the individual models are recalibrated and spot price predictions are made for February 12, 2011, etc. Consequently, the first combined forecast (using SIMPLE and LAD averaging) is made for February 11, 2011, the next for February 12, 2011, etc. Finally, the third date in Fig. 1—September 23, 2011—marks the first day for which PI are calculated. This is done using either (i) ARX and SNARX individual models and the empirical scheme for computing PI from the models' point forecasting errors, (ii) SIMPLE and LAD averaging and the empirical scheme for computing PI from the errors of the combined point forecasts or (iii) point forecasts of all 12 individual models and quantile regression, i.e. the new Quantile Regression Averaging (QRA) approach. The PI validation period lasts until January 14, 2012, and includes 114 days.

3 The techniques

3.1 Individual models

A typical and obvious feature shared by all empirical applications using forecast averaging is that results depend on the specific choice of individual models. Thus, the set of individual techniques considered here includes the 12 models analyzed by

Weron and Misiorek (2008) and then used in the context of averaging point forecasts by Nowotarski et al. (2014): autoregressive models (AR, ARX—the latter with temperature as the exogenous variable), spike preprocessed autoregressive models (p-AR, p-ARX; where the model structure was estimated after replacing price spikes with less extreme observations), threshold autoregressive models (TAR, TARX), mean-reverting jump diffusions (MRJD, MRJDX) and two classes of semiparametric autoregressive models (IHMAR, IHMARX, SNAR, SNARX; introduced by Weron and Misiorek 2008). In this study, we also use two of those individual models—ARX and SNARX—as benchmarks for comparison of the prediction intervals.

The ARX autoregressive model structure is given by the following formula:

$$p_t = \phi_1 p_{t-24} + \phi_2 p_{t-48} + \phi_3 p_{t-168} + \phi_4 m p_t + \psi_1 z_t + d_1 D_{Mon} + d_2 D_{Sat} + d_3 D_{Sun} + \varepsilon_t. \quad (1)$$

The lagged log-prices p_{t-24} , p_{t-48} and p_{t-168} account for the autoregressive effects of the previous days (the same hour yesterday, two days ago and one week ago), while $m p_t$ is the minimum of the previous day's 24 hourly log-prices and creates a link between bidding and price signals from the entire previous day (for discussions on the choice of this 'link' variable see Kristiansen 2012; Weron and Misiorek 2008). The variable z_t refers to the logarithm of the hourly temperature. The three dummy variables— D_{Mon} , D_{Sat} and D_{Sun} (for Monday, Saturday and Sunday, respectively)—account for the weekly seasonality. Finally, the ε_t 's are assumed to be independent and identically distributed (i.i.d.) with zero mean and finite variance. Model parameters are estimated in Matlab by minimizing the Final Prediction Error (FPE) criterion. Construction of empirical PI boils down to taking a desired quantile of the in-sample residuals, which resembles estimating Value-at-Risk via historical simulation, see Weron and Misiorek (2008).

The smoothed nonparametric ARX (or SNARX) model relaxes the normality assumption needed for the maximum likelihood estimation in the ARX model. It has the same functional form (1) but the parameter estimates are obtained from a numerical maximization of the empirical likelihood as suggested by Cao et al. (2003) and in the context of electricity price forecasting by Weron and Misiorek (2008). The empirical prediction intervals are constructed in the same way as for ARX. Note that SNARX was recommended by Weron and Misiorek as the best individual model out of the 12 considered specifications, mainly due to its good PI performance.

3.2 Combining point forecasts

The idea of combining forecasts goes back to the late 1960s, with the works of Bates and Granger (1969) and Crane and Crotty (1967). Since then, many authors have suggested the superior performance of forecast combinations over the use of individual models, see e.g. Clemen (1989), de Menezes et al. (2000), Timmermann (2006) and references therein. Despite this popularity, combining forecasts has not been discussed widely in the context of electricity spot prices. Only very recently Bordignon et al. (2013), Nowotarski et al. (2014) and Raviv et al. (2013) have provided empirical

support for the benefits of combining forecasts to obtain better point predictions of electricity spot prices.

In this paper we focus on computing prediction intervals derived from combined spot price forecasts and on evaluating their performance. To achieve this we examine two averaging schemes—SIMPLE averaging and least absolute deviation or LAD regression—that have been found by Nowotarski et al. (2014) to provide accurate and robust results. SIMPLE averaging is the most natural approach to combining forecasts and boils down to computing the (arithmetic) mean of all forecasts produced by the individual models. It is highly robust and is widely used in business and economic forecasting, see e.g. Clemen (1989) and Genre et al. (2013).

Another popular averaging method is based on classical linear regression. In this approach, the individual forecasts are regressors and the corresponding observed spot price is the dependent variable. The idea was first described by Crane and Crotty (1967), but it was the influential paper of Granger and Ramanathan (1984) to inspire more research effort in this direction. Since then, this way of averaging took on numerous variations according to different findings from different datasets. In this study we decide to follow an alternative proposed by Nowotarski et al. (2014) and replace the ordinary least squares approach with the absolute loss function. The resulting scheme is called least absolute deviation or LAD regression. An advantage of using the absolute loss function is its robustness to electricity price spikes. Indeed, a model that performs well in general, yet significantly underperforms on specific dates, is punished harder by the quadratic loss function. As a consequence it leads to a relatively large decrease of this model's weight, while using the absolute loss function yields a relatively smaller decrease of the weight.

The forecasting setup is the following. First, for each day t we compute $M = 12$ individual day-ahead price forecasts $\hat{p}_{1t}, \dots, \hat{p}_{Mt}$ of the 12 individual models (see Sect. 3.1). Then we combine them to yield a spot price forecast:

$$\hat{p}_t^c = \hat{\mathbf{p}}_t \mathbf{w}_t = \sum_{i=1}^M w_{it} \hat{p}_{it}, \quad (2)$$

where w_{it} is the weight assigned at time t to forecast i . We calculate the weights recursively at each time step, using data from the first prediction day (indicated by the first dotted vertical line in Fig. 1) until $t - 24$ (i.e. 24 h prior to the hour we forecast the price for). Like for individual forecasts, when predicting the price for hour h , we utilize past price values only for that particular hour, i.e. we split our data into 24 time series. As our main interest here are prediction intervals (not point forecasts), given weights at time t , we obtain the empirical prediction intervals by computing a quantile of the distribution of in-sample residuals of the averaged forecasts (2), i.e. analogously as for ARX and SNARX (see Sect. 3.1).

3.3 Quantile regression and forecast averaging

The quantile regression model introduced by Koenker and Bassett (1978) has been widely used in many financial and economic applications, out of which the most

intuitive may be Value-at-Risk as it simply concerns a quantile of a portfolio (Bunn et al. 2013). Quantile regression allows us to describe the conditional distribution of the dependent variable given its explanatory variables. Our idea here is to apply quantile regression to point (not interval!) forecasts of the individual models, i.e. to use the individual point forecasts as independent variables and the corresponding observed spot price as the dependent variable. Note that our method—dubbed Quantile Regression Averaging (QRA)—yields an interval forecast of the spot price, but does not use the PI of the individual methods. This is an important point, since as Wallis (2005) remarks: *combining intervals directly will not in general give an interval with the correct probability*. For instance, Granger et al. (1989) attempt to overcome this difficulty by estimating combining weights from data on past forecasts that in effect recalibrate the forecast quantiles, which is a cumbersome solution.

In our case the averaging problem is given by:

$$Q_p(q|\hat{\mathbf{p}}_t) = \hat{\mathbf{p}}_t \mathbf{w}_q, \quad (3)$$

where $Q_p(q|\cdot)$ is the conditional q th quantile of the electricity spot price distribution, $\hat{\mathbf{p}}_t$ are the regressors (explanatory variables) and \mathbf{w}_q is a vector of parameters (q in the subscript emphasizes the fact that the parameters are varying for different quantiles). The weights are estimated by minimizing the loss function for a particular q th quantile:

$$\begin{aligned} \min_{\mathbf{w}_t} & \left[\sum_{\{t: p_t \geq \hat{\mathbf{p}}_t \mathbf{w}_t\}} q |p_t - \hat{\mathbf{p}}_t \mathbf{w}_t| + \sum_{\{t: p_t < \hat{\mathbf{p}}_t \mathbf{w}_t\}} (1 - q) |p_t - \hat{\mathbf{p}}_t \mathbf{w}_t| \right] \\ & = \min_{\mathbf{w}_t} \left[\sum_t (q - \mathbb{1}_{p_t < \hat{\mathbf{p}}_t \mathbf{w}_t}) (p_t - \hat{\mathbf{p}}_t \mathbf{w}_t) \right]. \end{aligned} \quad (4)$$

Finally, note that the LAD regression may be viewed as a special case of quantile regression by taking the quantile to be the median, i.e. $q = 0.5$. Indeed, for $q = 0.5$ the right hand side of formula (4) under the minimum reduces to $\sum_t \frac{1}{2} |p_t - \hat{\mathbf{p}}_t \mathbf{w}_t|$ and, hence, we obtain the absolute loss function (up to a constant). That said, the proposed method is a natural extension of the averaging problem, where the point of interest is a point forecast.

It should be also emphasized that the QRA-based 50 % PI are not the same as the LAD-based 50 % PI. The former are based on running quantile regression for $q = 0.25$ and $q = 0.75$, the latter on running quantile regression for $q = 0.5$ and then taking the 25 and 75 % quantiles of the distribution of forecast errors (i.e. residuals).

4 Results

First, we evaluated the quality of the interval forecasts by comparing the nominal coverage to the true coverage. Thus, for the two individual models (ARX and SNARX) and for the three averaging schemes (SIMPLE, LAD and QRA) we calculated PIs

Table 1 *Upper part:* Unconditional coverage of the 50 and 90 % two-sided day-ahead PIs by the actual spot price for all considered models. *Lower part:* Descriptive statistics of the PI width

PI (%)	ARX	SNARX	SIMPLE	LAD	QRA
<i>Unconditional coverage</i>					
50	69.74	56.51	58.63	56.36	53.55
90	96.13	94.23	94.44	93.64	92.07
<i>Mean (standard deviation) of the PI width</i>					
50	8.63 (3.33)	6.09 (2.64)	6.32 (2.89)	6.73 (3.66)	6.40 (3.78)
90	21.28 (8.29)	20.73 (8.78)	25.73 (15.74)	26.20 (17.21)	21.10 (12.09)
<i>Median (inter-quartile range) of the PI width</i>					
50	8.66 (5.25)	5.94 (4.21)	5.89 (5.77)	5.79 (6.93)	5.62 (5.19)
90	21.34 (13.02)	20.64 (15.28)	23.22 (25.86)	21.87 (26.33)	19.51 (18.51)

See also Fig. 3 and Table 2

and determined the percentage of coverage of the 50 and 90 % two sided day-ahead PI by the actual spot price. If the model implied interval forecasts were accurate then the percentage of coverage should match the nominal values. For each model, 24×114 hourly values were determined and compared to the actual spot price. The unconditional coverage is summarized in the upper part of Table 1 and in Table 2. Interestingly, all models have overall a slightly higher coverage than nominal (for some hours the coverage is lower than nominal, though), possibly due to a less volatile spot price in the PI validation period as compared to the calibration period, see Fig. 1. For both analyzed levels (50 and 90 %) the best results were achieved by QRA, while the worst by ARX.

Next, we applied the approach of Christoffersen (1998) to test the unconditional and conditional coverage. This model independent approach is designed to overcome the clustering effect. The tests are carried out in the likelihood ratio (LR) framework. Three LR statistics are calculated: for unconditional coverage, independence and conditional coverage. The former two are distributed asymptotically as $\chi^2(1)$ and the latter as $\chi^2(2)$. If we condition on the first observation, then the conditional coverage LR test statistics is the sum of the other two. It should be noted that the independence test and hence the conditional coverage test is conducted only with respect to the first order dependency of exceedances. As Clements and Taylor (2003) show, the test can be easily modified to measure higher order dependency; see Maciejowska et al. (2014) for a sample application of this approach in the context of electricity price forecasting and Berkowitz et al. (2011) for a review of more complex tests. For simplicity of exposition we present here only the results of the conditional coverage test based on the first order test for independence, however, the conclusions are qualitatively the same if higher order dependency is tested.

The conditional and unconditional coverage LR statistics are plotted in Fig. 2. We repeat the methodology of Weron and Misiorek (2008) and conduct the tests separately for the 24 hourly time series. It would not make sense to compute the statistics jointly for all hours, since, by construction, the forecasts for consecutive

Table 2 Unconditional coverage (in percent) of the 50 and 90 % two-sided day-ahead PIs by the actual spot price for all five considered methods, independently for each of the 24 h

Hour	ARX	SNARX	SIMPLE	LAD	QRA	ARX	SNARX	SIMPLE	LAD	QRA
<i>Unconditional coverage, 50 % PI</i>					<i>Unconditional coverage, 90 % PI</i>					
1	73.68	56.14	53.51	56.14	59.65	94.74	92.11	92.11	90.35	89.47
2	72.81	57.02	58.77	55.26	61.40	93.86	92.11	91.23	90.35	92.11
3	70.18	59.65	58.77	53.51	58.77	96.49	93.86	92.98	92.98	92.98
4	72.81	62.28	56.14	53.51	57.89	98.25	96.49	92.11	89.47	91.23
5	68.42	58.77	54.39	49.12	52.63	95.61	92.11	86.84	85.96	92.11
6	49.12	37.72	38.60	34.21	45.61	89.47	83.33	75.44	80.70	88.60
7	52.63	42.98	32.46	35.96	44.74	91.23	92.11	86.84	85.09	85.09
8	64.91	52.63	43.86	44.74	44.74	95.61	94.74	91.23	88.60	86.84
9	73.68	57.89	54.39	57.89	54.39	96.49	93.86	92.11	92.98	92.11
10	71.05	61.40	60.53	60.53	54.39	96.49	94.74	97.37	97.37	91.23
11	78.07	63.16	73.68	71.93	69.30	99.12	96.49	100.00	100.00	97.37
12	78.07	57.02	76.32	70.18	50.88	100.00	97.37	100.00	96.49	87.72
13	77.19	65.79	78.07	64.91	50.88	99.12	97.37	100.00	97.37	93.86
14	76.32	58.77	82.46	66.67	49.12	98.25	98.25	100.00	99.12	93.86
15	81.58	71.05	79.82	58.77	48.25	99.12	98.25	100.00	99.12	93.86
16	78.95	64.04	74.56	74.56	62.28	98.25	97.37	100.00	100.00	99.12
17	62.28	49.12	49.12	46.49	42.98	92.98	92.98	97.37	97.37	93.86
18	59.65	57.89	50.00	50.00	51.75	93.86	93.86	93.86	92.98	92.98
19	53.51	42.98	45.61	55.26	50.88	92.98	91.23	93.86	94.74	91.23
20	65.79	52.63	57.89	57.02	55.26	96.49	95.61	97.37	95.61	92.98
21	72.81	58.77	63.16	67.54	60.53	98.25	96.49	98.25	97.37	93.86
22	75.44	53.51	55.26	63.16	57.02	98.25	95.61	96.49	96.49	95.61
23	72.81	53.51	52.63	50.00	48.25	96.49	93.86	97.37	93.86	89.47
24	71.93	61.40	57.02	55.26	53.51	95.61	91.23	93.86	92.98	92.11

See also Table 1

hours are correlated—predictions for all 24 h of the next day are made at the same time using the same information set.

According to Christoffersen's test statistics, QRA-based PI were the best among the tested approaches. Apart from two (four) peak hours, the QRA-based PI were not rejected by the conditional coverage (unconditional coverage) test at the 1 % level. For a majority of hours, they were not rejected at the 5 % level as well. On the other hand, the ARX model turned out to be the worst out of all analyzed methods according to the unconditional coverage test, with the null hypothesis rejected in 21 and 17 out of 24 cases, respectively for the 50 and 90 % PI at 5 % confidence level. Empirical prediction intervals derived from SIMPLE averaging yielded better results than ARX. However, for the hours from 11 to 16 the SIMPLE averaging implied PI were generally too wide, see also the lower part of Table 1 and Fig. 3. As a result the test statistics were in excess of 20 and the null hypotheses were rejected for the two analyzed PI

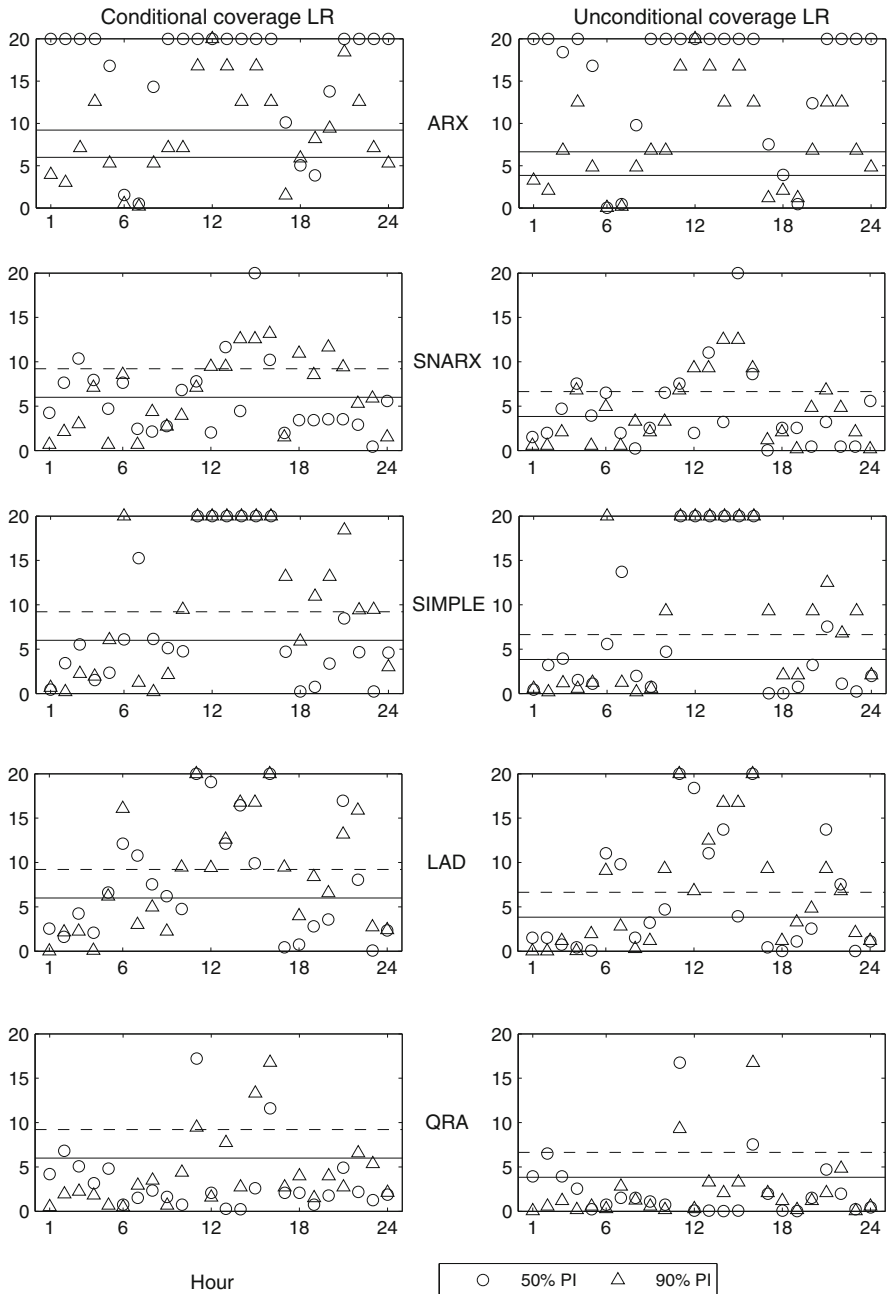


Fig. 2 The conditional coverage (left panels) and unconditional coverage (right panels) LR statistics for the PI obtained from the five considered forecasting models. The solid (dashed) horizontal lines represent the 5% (1%) significance level of the appropriate χ^2 distribution. All test values exceeding 20 are set to 20

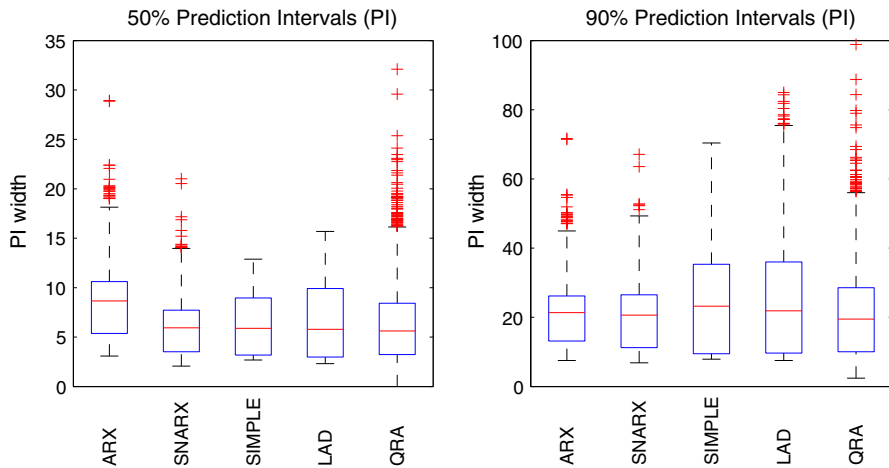


Fig. 3 Boxplots of widths of the prediction intervals (PI) for the five considered forecasting models

coverages (50 and 90 %). Moreover, for those hours SIMPLE based 90 % PI had a 100 % coverage. For the rest of the day the method performed slightly better, especially during night and early morning hours. The LAD regression did not perform much better. The test statistics were similar to those achieved by SIMPLE and better than the benchmark ARX.

Finally note that the best individual method according to [Weron and Misiorek \(2008\)](#)—SNARX—performed worse than QRA, but better than the remaining three approaches. The SNARX based PI were on average narrower and less volatile than QRA-based, but their median was higher than that of QRA-based PI, see [Table 1](#). It seems as if QRA yields PI which adapt themselves better to the changing price dynamics than those obtained from the SNARX model.

5 Conclusions

This paper examines possible accuracy gains from forecast averaging in the context of interval forecasts of electricity spot prices. While there is a significant number of studies on the use of forecast combinations for constructing interval forecasts of economic and financial variables, to our best knowledge, there are no publications where this approach would be tested on the extremely volatile electricity spot price data. Our paper can be considered as an extension of the empirical studies of [Weron and Misiorek \(2008\)](#) and [Nowotarski et al. \(2014\)](#). In the former article the same individual models were evaluated in a point and interval forecasting exercise. The smoothed nonparametric autoregressive SNARX model came out as the best in terms of PIs. In the latter paper the same individual models were used as the building blocks for eight approaches to combining point forecasts. The SIMPLE and LAD averaging schemes turned out to be relatively accurate and robust ways to combine electricity spot price forecasts.

In this study we (i) construct empirical PI from combined electricity spot price forecasts, utilizing SIMPLE and LAD averaging, and (ii) propose a new method—

Quantile Regression Averaging (QRA)—for constructing prediction intervals using the concept of quantile regression and a pool of point forecasts of individual (i.e. not combined) time series models. It turns out that the empirical PI from combined forecasts outperform ARX model implied PI, in terms of both the unconditional and conditional coverage, but are outperformed by the SNARX based PI. However, the QRA-based PI are found to be more accurate than those of the semiparametric SNARX model. The latter fact is particularly appealing since SNARX yielded the most accurate PI of out the 12 individual models used in this study. In other words, the Quantile Regression Averaging approach we propose here—which is obviously dependent on the quality of the point forecasts of the individual models—outperformed the best individual model.

Acknowledgments We are grateful to Eran Raviv and Stefan Trück for inspiring discussions on forecast averaging and electricity spot price forecasting, and to two anonymous reviewers for their critical comments. This paper has also benefited from conversations with the participants of the Applicable Semiparametrics Conference (2013), the European Energy Market Conferences (EEM13, EEM14), the Conference on Energy Finance (EF2013) and the Energy Finance Christmas Workshop (EFC13). This work was supported by funds from the National Science Centre (NCN, Poland) through Grants Nos. 2011/01/B/HS4/01077 and 2013/11/N/HS4/03649.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Aggarwal SK, Saini LM, Kumar A (2009) Load/price forecasting and managing demand response for smart grids. *Electr Power Energy Syst* 31:13–22
- Bates JM, Granger CW (1969) The combination of forecasts. *Oper Res Q* 20:451–468
- Berkowitz J, Christoffersen P, Pelletier D (2011) Evaluating Value-at-Risk models with desk-level data. *Manage Sci* 57(12):2213–2227
- Bierbrauer M, Menn C, Rachev ST, Trück S (2007) Spot and derivative pricing in the EEX power market. *J Bank Finance* 31:3462–3485
- Bordignon S, Bunn DW, Lisi F, Nan F (2013) Combining day-ahead forecasts for British electricity prices. *Energy Econ* 35:88–103
- Bunn D, Andresen A, Chen D, Westgaard S (2013) Analysis and forecasting of electricity price risks with quantile factor models. London Business School Working Paper. Available from http://www.ceem-dauphine.org/assets/dropbox/Derek_BUNN.pdf
- Cao R, Hart JD, Saavedra A (2003) Nonparametric maximum likelihood estimators for ar and ma time series. *J Stat Comput Simul* 73:347–360
- Chatfield C (ed.) (2000) *Time Ser Forecast*. Chapman & Hall/CRC, Boca Raton
- Christoffersen P (1998) Evaluating interval forecasts. *Int Econ Rev* 39(4):841–862
- Clemen RT (1989) Combining forecasts: a review and annotated bibliography. *Int J Forecast* 5:559–583
- Clements MP, Taylor N (2003) Evaluating interval forecasts of high-frequency financial data. *J Appl Econom* 18:445–456
- Conejo AJ, Contreras J, Espínola R, Plazas MA (2005) Forecasting electricity prices for a day-ahead pool-based electric energy market. *Int J Forecast* 21:435–462
- Crane DB, Crotty JR (1967) A two-stage forecasting model: exponential smoothing and multiple regression. *Manage Sci* 6:B501–B507
- de Menezes LM, Bunn DW, Taylor JW (2000) Review of guidelines for the use of combined forecasts. *Eur J Oper Res* 120:190–204
- Eydeland A, Wolyniec K (2003) *Energy and power risk management*. 2nd edn. Wiley, Hoboken

- Garcia-Martos C, Rodriguez J, Sanchez M (2012) Forecasting electricity prices by extracting dynamic common factors: application to the Iberian market. *IET Gener Transm Distrib* 6:11–20
- Genre V, Kenny G, Meyler A, Timmermann A (2013) Combining expert forecasts: can anything beat the simple average? *Int J Forecast* 29(1):108–121
- Granger C, Ramanathan R (1984) Improved methods of combining forecasts. *J Forecast* 3:197–204
- Granger C, White H, Kamstra M (1989) Interval forecasting: an analysis based upon arch-quantile estimators. *J Econom* 40:87–96
- Kaminski V (2013) *Energy markets*. Risk Books, New York
- Karakatsani N, Bunn D (2008) Forecasting electricity prices: the impact of fundamentals and time-varying coefficients. *Int J Forecast* 24:764–785
- Koenker R, Basset G (1978) Regression quantiles. *Econometrica* 46(1):33–50
- Kristiansen T (2012) Forecasting Nord Pool day-ahead prices with an autoregressive model. *Energy Policy* 49:328–332
- Maciejowska K, Nowotarski J, Weron R (2014) Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging. *Int J Forecast* (submitted). Working paper version available from RePEc: <http://ideas.repec.org/p/wuu/wpaper/hsc1409.html>
- Maciejowska K, Weron R (2013) Forecasting of daily electricity spot prices by incorporating intra-day relationships: evidence from the UK power market. In: *IEEE Conference proceedings, 10th international conference on the European energy market (EEM'13)*, 28–30 May 2013, Stockholm, Sweden. doi:10.1109/EEM.2013.6607314
- Misiorek A, Trück S, Weron R (2006) Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models. *Stud Nonlinear Dyn Econ* 10(3), Article 2. http://www.efm.leeds.ac.uk/~mark/ISIabbr/S_abrvjt.html
- Nogales FJ, Conejo AJ (2006) Electricity price forecasting through transfer function models. *J Oper Res Soc* 57:350–356
- Nowotarski J, Raviv E, Trück S, Weron R (2014) An empirical comparison of alternative schemes for combining electricity spot price forecasts. *Energy Econ*. doi:10.1016/j.eneco.2014.07.014
- Peña JI (2012) A note on panel hourly electricity prices. *J Energy Mark* 5:81–97
- Raviv E, Bouwman KE, van Dijk D (2013) Forecasting day-ahead electricity prices: Utilizing hourly prices. Tinbergen Institute Discussion Paper 13-068/III. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.2266312>.
- Timmermann AG (2006) Forecast combinations. In: Elliott G, Granger CW, Timmermann A (eds) *Handbook of economic forecasting*. Elsevier, Amsterdam, pp 135–196
- Wallis KF (2005) Combining density and interval forecasts: a modest proposal. *Oxf Bull Econ Stat* 67:983–994
- Weron R (2006) *Modeling and forecasting electricity loads and prices: a statistical approach*. Wiley, Chichester
- Weron R (2014) Electricity price forecasting: a review of the state-of-the-art with a look into the future. *Int J Forecast*. doi:10.1016/j.ijforecast.2014.08.008
- Weron R, Misiorek A (2008) Forecasting spot electricity prices: a comparison of parametric and semiparametric time series models. *Int J Forecast* 24:744–763
- Wu H, Chan S, Tsui K, Hou Y (2013) A new recursive dynamic factor analysis for point and interval forecast of electricity price. *IEEE Trans Power Syst* 28(3):2352–2365